NOAA Technical Memorandum NMFS

SAMPLE SIZE RECOMMENDATIONS FOR ESTIMATING STOCK COMPOSITION USING GENETIC STOCK IDENTIFICATION (GSI)

Shanae D. Allen-Moran[1,2]
William H. Satterthwaite[1]
Michael S. Mohr[1]

[1] National Marine Fisheries Service
Southwest Fisheries Science Center
Fishery Ecology Division
110 Shaffer Rd.
Santa Cruz, CA 95060

[2] Institute of Marine Sciences
University of California, Santa Cruz
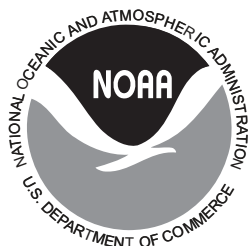1156 High Street
Santa Cruz, CA 95064

U. S. DEPARTMENT OF COMMERCE
National Oceanic and Atmospheric Administration
National Marine Fisheries Service
Southwest Fisheries Science Center

The National Oceanic and Atmospheric Administration (NOAA), organized in 1970, has evolved into an agency that establishes national policies and manages and conserves our oceanic, coastal, and atmospheric resources. An organizational element within NOAA, the Office of Fisheries, is responsible for fisheries policy and the direction of the National Marine Fisheries Service (NMFS).

In addition to its formal publications, the NMFS uses the NOAA Technical Memorandum series to issue informal scientific and technical publications when complete formal review and editorial processing are not appropriate or feasible. Documents within this series, however, reflect sound professional work and may be referenced in the formal scientific and technical literature.

JUNE 2013

SAMPLE SIZE RECOMMENDATIONS FOR ESTIMATING STOCK COMPOSITION USING GENETIC STOCK IDENTIFICATION (GSI)

Shanae D. Allen-Moran[1,2]
William H. Satterthwaite[1]
Michael S. Mohr[1]

[1] National Marine Fisheries Service
Southwest Fisheries Science Center
Fishery Ecology Division
110 Shaffer Rd.
Santa Cruz, CA 95060

[2] Institute of Marine Sciences
University of California, Santa Cruz
1156 High Street
Santa Cruz, CA 95064

NOAA-TM-NMFS-SWFSC-513

1 Abstract

Genetic Stock Identification (GSI) allows estimation of stock composition, including untagged stocks. The required sample size per sampling stratum depends on the commonness/rarity of the stock(s) of interest and, when the stock composition of the landings itself is of interest rather than the ocean mixed stock aggregate, on total catch as well. We present sample sizes required to achieve specified precisions, detect rare stocks, and test the significance of differences between proportions. For the ocean mixed stock aggregate, sampling 800 fish per stratum suffices to estimate stock proportions $\geq 3\%$ with a coefficient of variation $\leq 20\%$ and has a high ($> 99.9\%$) probability of detecting stocks comprising $\geq 1\%$ of the mixed stock, but this sample size does not yield precise estimates of small stock proportions and may fail to detect rare stocks. To precisely estimate smaller proportions, required sample size scales approximately inversely with the proportion, i.e., halving the target proportion requires doubling the sample size. For a given proportion, increased precision also comes at a greater cost as the required sample size scales approximately with the square of the desired precision, i.e., halving the target standard error requires quadrupling the sample size. Smaller sample sizes suffice for estimating the stock composition of the landings itself when total catch is low, but a high sampling fraction may be required. Statistical comparisons between two proportions require larger sample sizes than estimating single proportions of similar magnitude.

2 Introduction

West coast salmon fisheries management requires understanding spatial and temporal patterns of stock-specific distribution and fisheries impacts. To date, information used in management has primarily come from the analysis of recoveries of coded-wire tags (CWT, Johnson 1990; OSP 1999). In addition to their use in management models (e.g., Goldwasser et al. 2001; O'Farrell et al. 2012), CWT data have provided insight into stock-specific ocean distributions (Weitkamp and Neely 2002; Trudel et al. 2009; Weitkamp 2010), and the availability of multi-decade datasets has allowed quantifying temporal variability in size-at-age (Satterthwaite et al. 2012) and spatial distribution (Satterthwaite et al. 2013). However, the CWT methodology is limiting in that only very rarely are fish tagged outside of hatcheries, and for some stocks of interest there is no tagging program at all. In addition, recovering a CWT from a live tagged fish is lethal, which may preclude its use in determining the stock composition of fishery discards, for example. Currently, approximately 20% of all west coast ocean salmon landings are examined for the recovery of CWTs, and to obtain estimates of harvest and fishing effort stratified by time, area, and fishery (Johnson 1990; OSP 1999).

Genetic stock identification (GSI) methods have now been sufficiently developed (Seeb et al. 2007) to consider their use as a complement to the CWT program. GSI techniques rely on genetic markers, such as microsatellites or single-nucleotide polymorphisms, to probabilistically assign each sampled fish to a baseline set of potential genetic stock units (Seeb et al. 2007). The resolution of GSI thus includes the untagged stocks, as long as they are individually represented in the set of baseline stocks. In addition, with GSI every fish sampled is potentially informative with respect to stock composition, not just tagged individuals. Thus, adequate sample sizes might be possible using fishery-independent sampling methods. Finally, collecting tissue from a live sampled fish for genotyping purposes is non-lethal, although released sample fish are subject to an associated release mortality rate estimated to be 26% and 19% for west coast ocean salmon commercial and recreational fishing methods, respectively (STT 2000).

While GSI has some notable benefits and addresses certain topics that CWT methods cannot,

there are also drawbacks to the use of GSI for estimation of stock composition. In particular, the resolution of the current set of baseline genetic stock units ("reporting groups") does not allow for unambiguous discrimination of several stocks that are important for west coast fisheries management. With respect to California stocks, for example, Klamath River fall and spring run Chinook salmon cannot be distinguished, nor can Sacramento River fall, Sacramento River late-fall, San Joaquin fall, and Feather River Hatchery spring run Chinook salmon (Seeb et al. 2007). Second, there is the possibility of stock assignment errors, and of individual samples that either do not assign to any stock with high probability, or are unusable due to problems with tissue collection and preservation. Third, GSI alone cannot provide the age of the sampled fish, which is essential for many fishery assessments.

The greatest value of GSI data may lie in providing insight about untagged stocks for which no other information on ocean harvest is available. Since most highly abundant stocks are supplemented by hatchery production with tagging programs, untagged stocks are typically lower-abundance stocks that make up small proportions of the ocean harvest. Thus to realize their full potential, GSI sampling programs must use adequate sample sizes for acceptable statistical performance when estimating small proportions.

GSI sample size recommendations are developed in this paper for the estimation of stock composition of the ocean mixed stock aggregate and of fishery landings based on a clearly defined set of potential statistical objectives. These objectives include:

1. The probability of detecting at least k individuals of a given stock meets a specified criterion.

2. The coefficient of variation of an estimated stock proportion meets a specified criterion.

3. The standard error of an estimated stock proportion meets a specified criterion.

4. The power to detect differences between two stock proportions meets a specified criterion.

The set of statistical objectives considered is defined in terms of a particular stock and sampling stratum of interest. The objectives are similar to those proposed by the Pacific Salmon Commission

3

GSI modelling and sampling workgroup (Mohr 2007); however, we additionally consider them in the context of sampling without replacement. For a given application it may be desirable to achieve one or more of these objectives, or it may be desirable to consider the entire set of objectives in designing a broad-based sampling plan.

Objective 1, evaluated for $k = 1$, concerns the ability to detect a stock given its presence, but larger values of k may be of interest for analyses of stock-specific characteristics requiring a minimum number of individuals. Objectives 2 and 3 both involve the precision of stock proportion estimates (and thus for fishery landings the precision of catch estimates, given known total catch), however Objective 2 considers error in relative terms whereas Objective 3 does so in absolute terms. Thus, Objective 2 may be more relevant for rare stocks. Objective 4 is relevant to the statistical comparison of proportions between strata or substrata for a given stock, or between stocks within a stratum.

The sample size necessary to satisfy these objectives depends on the magnitude of the stock proportion of interest and, for estimating the stock composition of landings, the stratum total catch. It is assumed here that all fish sampled for GSI are correctly identified to their stock of origin. To the extent that some samples cannot be successfully processed or genotyped, or are misclassified, the necessary sample size will be greater than that indicated by our results.

3 Methods

Denote by n the sample size, by Y the number of the stock of interest in the sample, and by $\hat{p} = \frac{Y}{n}$ the sample estimate of the stock proportion of interest, p (see Table 1 for the notation used in this paper). To determine the n necessary to satisfy the objectives, it is further necessary to specify the sampling distribution of Y that the estimated proportion is based on. The sampling distribution of Y in turn depends on whether the sampling is with or without replacement, and secondly what type of selection scheme is used in the with or without replacement sampling design. Regarding the latter, for the purposes of this paper we assume that the selection scheme used is "simple random"

Table 1. Notation used in this paper.

Symbol	Description
α	type-I error rate
β	type-II error rate; $1 - \beta$ is the statistical power
$CV_{\hat{p}}$	coefficient of variation of $\hat{p}$
n	sample size
N	total landings
p	proportion of stock of interest
$\hat{p}$	sample estimate of p
Q_k	probability that at least k individuals from stock of interest are sampled
$\sigma_{\hat{p}}^2$	sampling variance of $\hat{p}$
Y	number of stock of interest in sample of size n
Z_r	standard normal distribution r^{th} quantile (e.g., $Z_{0.975} = 1.96$)

(all individuals have an equal probability of being included in the sample) in both the with and without replacement designs.

Whether to assume with or without replacement sampling for the purpose of sample size planning depends, in part, on what the frame of reference is for the stock proportion estimate. Is it the stock proportion in the ocean mixed stock aggregate for which the estimate is desired, or is it the stock proportion in the landings itself? For the former, the sampling scheme can effectively be considered as with replacement sampling, regardless of whether the GSI sample is taken from fishery landings or through a fishery-independent scientific survey, since in both instances the number of fish sampled for GSI will likely be less than 5% of the ocean mixed stock aggregate.[1] In this case, the sampling distribution of Y is *binomial*(n, p), $\hat{p}$ is an unbiased estimator of p, and

---

[1]Of course, strictly speaking this does constitute without replacement sampling assuming that a fish sampled is not released and subject to resampling. However, at sampling fractions less than, say 5%, the statistical properties of without replacement sampling are very closely approximated by those of with replacement sampling. Regardless, application of the hypergeometric distribution in this situation would require specifying the size of the stratum-specific ocean population being sampled, which will of course be unknown.

binomial

$$\text{Prob}(Y = y) = \binom{n}{y} p^y (1-p)^{n-y}, \qquad \text{for } 0 \leq y \leq n, \quad 0 \leq p \leq 1, \tag{1}$$

$$\sigma_{\hat{p}}^2 = \frac{p(1-p)}{n}, \tag{2}$$

for integer-valued Y (Johnson et al. 1993; Cochran 1977). On the other hand, if the target of estimation is limited to the stock composition of the landings itself, the sampling fraction may exceed 5% in which case the sampling distribution of Y is $hypergeometric(N, n, p)$, $\hat{p}$ is an unbiased estimator of p, and

hypergeometric

$$\text{Prob}(Y = y) = \frac{\binom{Np}{y} \binom{N(1-p)}{n-y}}{\binom{N}{n}}, \qquad \text{for } \max(0, n - N(1-p)) \leq y \leq \min(n, Np), \tag{3}$$

$$\text{for } p = \left(\frac{0}{N}, \frac{1}{N}, \frac{2}{N}, \ldots, \frac{N}{N} \right),$$

$$\sigma_{\hat{p}}^2 = \frac{p(1-p)}{n} \cdot \frac{N-n}{N-1}, \tag{4}$$

for integer-valued Y and N (Johnson et al. 1993; Cochran 1977).

Note that a GSI sample taken from the landings might be used for either purpose, and while the estimated proportion $\hat{p}$ would be the same in either case, the sampling variance $\sigma_{\hat{p}}^2$ would be less for the without replacement scheme. The sampling variance is reduced in both schemes by increasing n, but in without replacement sampling $\sigma_{\hat{p}}^2$ is further reduced by an increased sampling fraction, $\frac{n}{N}$. In the limit, if all of the landings were sampled, the landing composition would be known exactly, but the composition of the ocean mixed stock aggregate would not.

The formulas above make clear that $\sigma_{\hat{p}}^2$ depends on n and p. In the case of without replacement sampling it also depends on N, but primarily through the sampling fraction: $\frac{N-n}{N-1} \approx 1 - \frac{n}{N}$. Thus, for small $\frac{n}{N}$ the formulas are nearly equivalent. For both distributions, the maximum of $\sigma_{\hat{p}}^2$ given n and N occurs when $p = 0.5$, and $\sigma_{\hat{p}}^2$ approaches 0 as p approaches 0 or 1.

For sample size planning purposes, the formulas above may be rearranged to give n as a func-

tion of $\text{Prob}(Y = y)$ or $\sigma_{\hat{p}}^2$ for a specific value of p and N (if applicable), permitting evaluation of Objectives 1–4. We do this for with replacement (binomial) and without replacement (hypergeometric) sampling by evaluating the smallest value of n necessary to meet the objectives over a range of p. We focused primarily on values of $p \leq 0.05$ because this likely characterizes most Chinook salmon stocks found off California (Winans et al. 2001), and because the level of n necessary to satisfy all but one of the considered objectives increases with decreasing p. For p smaller than 0.005, the necessary n increases steeply for several of the objectives, limiting the usefulness of our graphical presentation of the results of applying these formulas for n. Nevertheless, stock proportions less than 0.005 may indeed be of interest, e.g., when evaluating rare stocks, or stocks on the periphery of their geographic range.

For hypergeometric sampling, we evaluate sample size requirements over the range of p for $N = 100, 500$, or 1000, noting that with larger N the results approach those for binomial sampling. With hypergeometric sampling, p is a discrete-valued variable (equation (3)). This discreteness of p, and hence of the metric being evaluated, will be evident in the graphical presentation of these results, most notably for the $N = 100$ figure panels where the plotted range of $p \leq 0.05$ consists of only five non-zero values, $p = (0.01, 0.02, 0.03, 0.04, 0.05)$.

4   Results

4.1   Objective 1

The first objective is that n be sufficient to ensure, with probability Q_k, that the sample will contain at least k fish of the stock of interest. That is, $Q_k = \text{Prob}(Y \geq k) = 1 - \text{Prob}(Y < k)$. In this section, we evaluate this for the case of simple detection ($k = 1$); larger values of k are evaluated in Appendix A. For binomial sampling with $k = 1$, appealing to equation (1)

$$Q_1 = 1 - \text{Prob}(Y = 0) = 1 - (1-p)^n, \tag{5}$$

and solving for n gives

$$n = \frac{log(1-Q_1)}{log(1-p)}.$$ (6)

For hypergeometric sampling with $k = 1$, appealing to equation (3)

$$Q_1 = 1 - \text{Prob}(Y = 0) = 1 - \frac{\binom{N(1-p)}{n}}{\binom{N}{n}},$$ (7)

from which n can be approximated as (Fosgate 2009)

$$n = \left(1 - (1-Q_1)^{\frac{1}{Np}}\right)\left(N - \frac{Np-1}{2}\right).$$ (8)

An exact solution for n can be obtained as described in Appendix A, however, over the range of N, p, and Q_1 examined in Figure 1, the exact value differed from the approximation by no more than 1.

Figure 1 shows the n necessary to yield a specified Q_1, depending on p, under both binomial sampling and hypergeometric sampling for select values of N. Mohr (2007) suggests that Q_1 be at least 0.999 for all $p \geq 0.01$. Since for any particular value of Q_1 the required n increases as p approaches 0, the minimum n that will meet this condition is that which satisfies an Objective 1 criterion of $Q_1 = 0.999$ for $p = 0.01$. This is achieved with an n of 688 under binomial sampling, 497 for hypergeometric sampling with $N = 1000$ (50% sampling), 373 for $N = 500$ (75% sampling), and 100 for $N = 100$ (100% sampling). For larger N, the hypergeometric results approach the binomial results.

Note that in equation (6), if Q_1 is held constant, the required n scales with $\frac{-1}{log(1-p)}$ (negative since the numerator is the log of a quantity less than 1). For "small" p, $log(1-p) \approx -p$ (first term of Taylor expansion). Thus the required n for binomial sampling scales with $\frac{1}{p}$, i.e., halving the target proportion requires doubling n. Sample sizes can be approximated quite well via this method for $p < 0.01$, e.g., for $Q_1 = 0.999$ the required n for $p = 0.01$ is 688 while the required n for $p = 0.005$ is 1379.
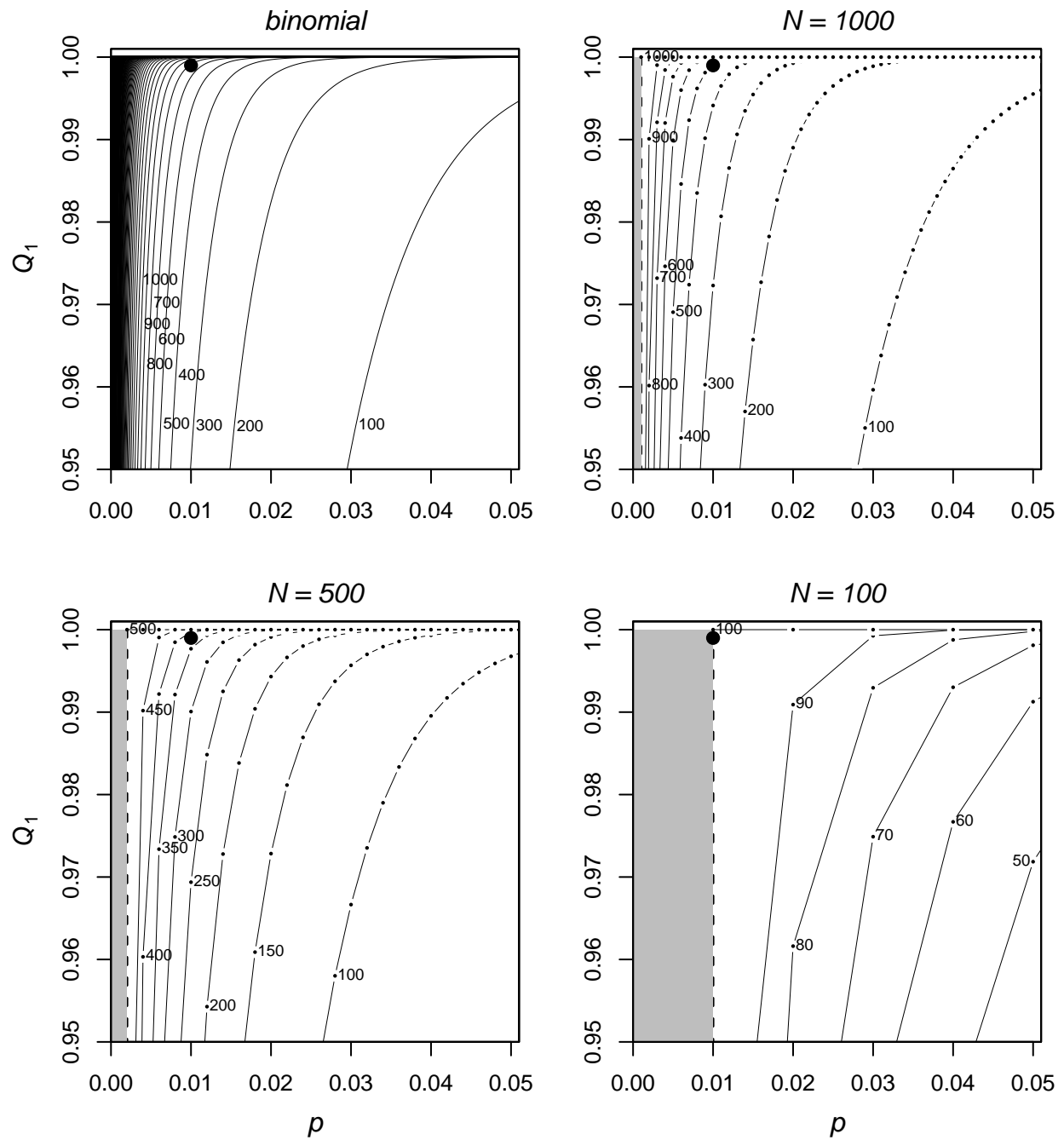
Figure 1. Sample size (contours) required to ensure, with probability Q_1, that the sample will contain at least one fish of the stock of interest, given the stock proportion p. Top-left panel presents results for binomial sampling; remaining panels give hypergeometric results for total landings of 1000, 500, and 100 fish, respectively. In the hypergeometric panels, p less than the minimum value possible $\left(\frac{1}{N}\right)$ for Q_1 are block-shaded in grey, and for contoured values of n the possible (n, p) combinations appear as black dots connected by line segments. Large dot marks criterion described in the text ($Q_1 = 0.999$ for $p = 0.01$).

Under binomial sampling, for a fixed $n \geq 30$, a decrease in Q_1 from 0.999 to 0.95 results in a 55% decrease in the minimum estimable proportion, while a change in Q_1 from 0.999 to 0.99 results in a 32% decrease. Conversely, as Q_1 surpasses 0.999, the minimum proportion that can be estimated for a given sample size increases quickly. For $p > 0.02$, sample sizes as small as 200 satisfy this objective over a wide range of Q_1 values (Figure 1).

For $k > 1$, the required n increases as described in Appendix A.

4.2 Objective 2

The second objective is that n be sufficient to ensure that the coefficient of variation of $\hat{p}$,

$$CV_{\hat{p}} = \frac{\sigma_{\hat{p}}}{p}, \tag{9}$$

a measure of relative precision, does not exceed a specified level. That is, the standard error of $\hat{p}$ should not exceed $(100 \cdot CV_{\hat{p}})\%$ of p. For binomial sampling, substituting the square root of equation (2) into equation (9) and solving for n gives

$$n = \frac{1-p}{p \cdot CV_{\hat{p}}^2}. \tag{10}$$

For hypergeometric sampling, substituting the square root of equation (4) into equation (9) and solving for n gives

$$n = \frac{N}{1 + CV_{\hat{p}}^2 \left(\frac{(N-1)p}{1-p} \right)}. \tag{11}$$

Figure 2 shows the n necessary to yield a specified $CV_{\hat{p}}$, depending on p, under both binomial sampling and hypergeometric sampling for select values of N. Mohr (2007) suggests that $CV_{\hat{p}}$ be at most 0.2 for all $p \geq 0.03$. Since for any particular value of $CV_{\hat{p}}$ the required n increases as p approaches 0, the minimum n that will meet this condition is that which satisfies an Objective 2 criterion of $CV_{\hat{p}} = 0.2$ for $p = 0.03$. This is achieved with an n of 809 under binomial sampling, 448 for hypergeometric sampling with $N = 1000$ (45% sampling), 310 for $N = 500$ (62% sam-
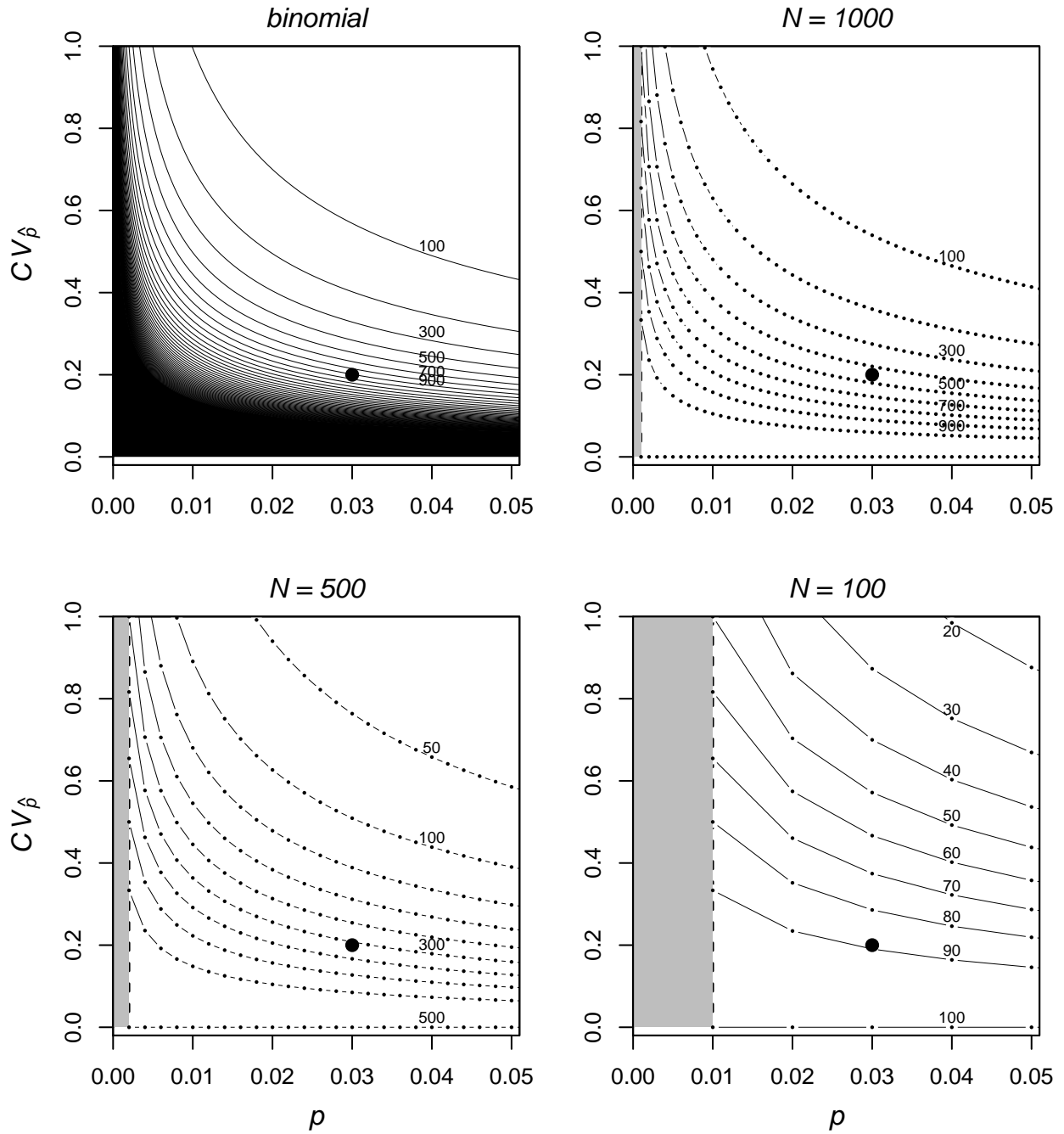
Figure 2. Sample size (contours) required to yield a specified $CV_{\hat{p}}$, given the stock proportion p. Top-left panel presents results for binomial sampling; remaining panels give hypergeometric results for total landings of 1000, 500, and 100 fish, respectively. In the hypergeometric panels, p less than the minimum non-zero value possible $\left(\frac{1}{N}\right)$ are block-shaded in grey, and for contoured values of n the possible (n, p) combinations appear as black dots connected by line segments. Large dot marks criterion described in the text ($CV_{\hat{p}} = 0.2$ for $p = 0.03$).

pling), and 90 for $N = 100$ (90% sampling). For larger N, the hypergeometric results approach the binomial results.

Note that for "small" p, equation (10) reduces to approximately $\frac{1}{p \cdot CV_{\hat{p}}^2}$ and thus, as with Objective 1, for a fixed $CV_{\hat{p}}$, the required n for binomial sampling scales with $\frac{1}{p}$, i.e., halving the target proportion requires a doubling of n. Alternatively, for a fixed p, the required n scales with $\frac{1}{CV_{\hat{p}}^2}$, i.e., halving the target $CV_{\hat{p}}$ requires a *quadrupling* of n.

4.3   Objective 3

The third objective is that n be sufficient to ensure that the standard error of $\hat{p}$, $\sigma_{\hat{p}}$, a measure of absolute precision, does not exceed a specified level. Equivalently, assuming $\hat{p}$ is normally distributed, this n will be sufficient to ensure that the 95% confidence interval width for p does not exceed $2 \cdot Z_{0.975} \cdot \sigma_{\hat{p}}$, or approximately $4 \cdot \sigma_{\hat{p}}$. For binomial sampling, solving equation (2) for n gives

$$n = \frac{p(1-p)}{\sigma_{\hat{p}}^2}. \tag{12}$$

For hypergeometric sampling, solving equation (4) for n gives

$$n = \frac{p(1-p)N}{\sigma_{\hat{p}}^2(N-1) + p(1-p)}. \tag{13}$$

Figure 3 shows the n necessary to yield a specified $\sigma_{\hat{p}}$, depending on p, under both binomial sampling and hypergeometric sampling for select values of N. Mohr (2007) suggests that $\sigma_{\hat{p}}$ be at most 0.015 (which corresponds approximately to a 95% confidence interval width of 0.06) for all $p \geq 0.01$. Since for any particular value of $\sigma_{\hat{p}}$ the maximum n required occurs with $p = 0.5$, the minimum n that will meet this condition is that which satisfies an Objective 3 criterion of $\sigma_{\hat{p}} = 0.015$ for $p = 0.5$. This is achieved with an n of 1112 under binomial sampling, 527 for hypergeometric sampling with $N = 1000$ (53% sampling), 346 for $N = 500$ (69% sampling), and 92 for $N = 100$ (92% sampling). For larger N, the hypergeometric results approach the binomial results.
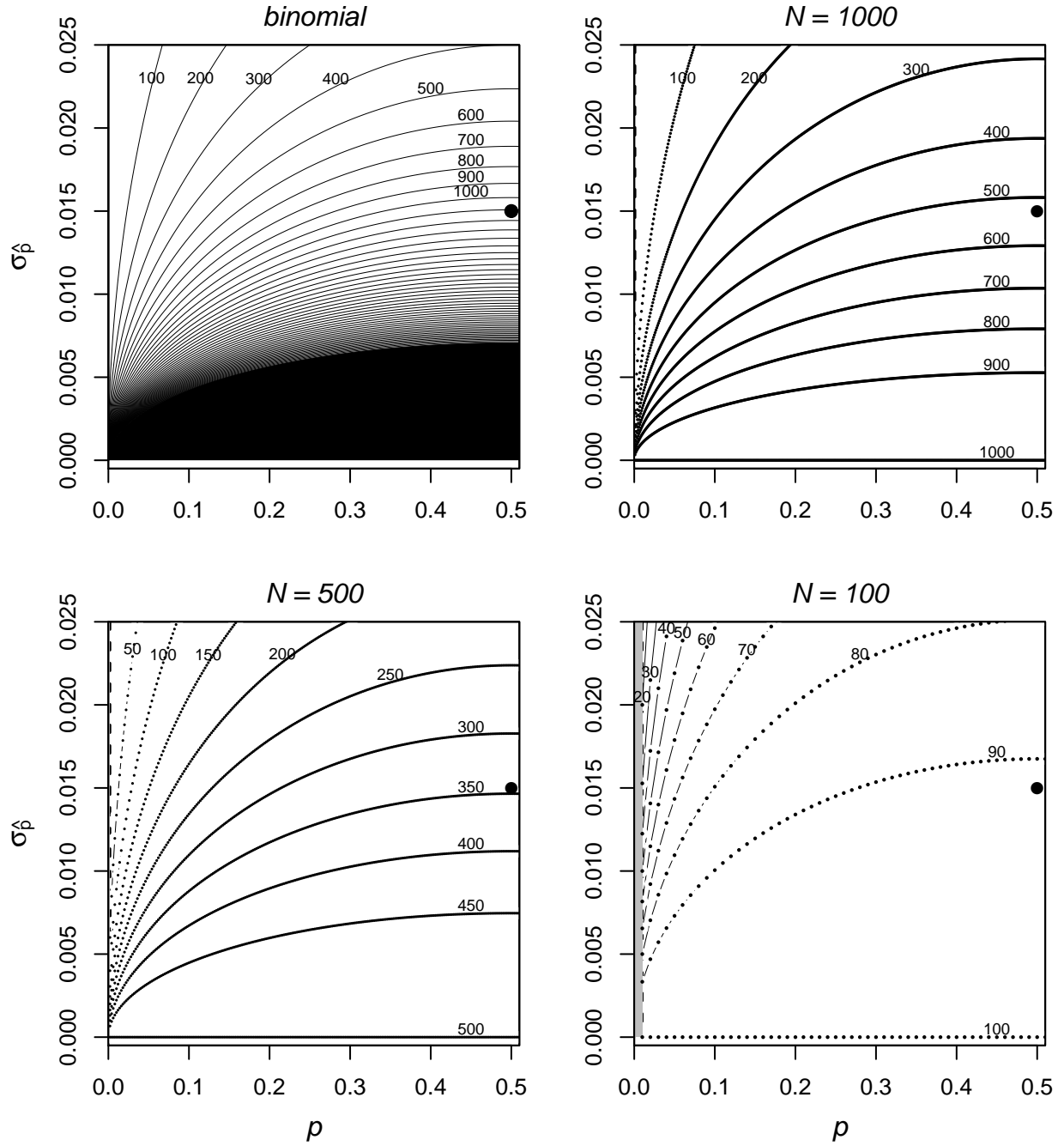
Figure 3. Sample size (contours) required to yield a specified $\sigma_{\hat{p}}$, given the stock proportion p. Top-left panel presents results for binomial sampling; remaining panels give hypergeometric results for total landings of 1000, 500, and 100 fish, respectively. In the hypergeometric panels, p less than the minimum non-zero value possible $\left(\frac{1}{N}\right)$ are block-shaded in grey, and for contoured values of n the possible (n, p) combinations appear as black dots connected by line segments (for larger N, n appearing as a nearly continuous line over the range displayed, $0 \leq p \leq 0.5$). Large dot marks criterion described in the text ($\sigma_{\hat{p}} = 0.015$ for $p = 0.5$).

13

Note that in equation (12) the required n for fixed p scales with $\frac{1}{\sigma_{\hat{p}}^2}$, so, like Objective 2, doubling the precision (halving $\sigma_{\hat{p}}$) requires quadrupling n. Unlike the other objectives considered in this paper, for a given $\sigma_{\hat{p}}$, n decreases as p approaches 0; although not shown, n also decreases as p approaches 1. Thus, small (or large) proportions are actually easier to estimate under an absolute precision criterion, than are proportions near 0.5. However, an acceptable standard error on a large proportion may be unacceptably large for a small proportion.

4.4 Objective 4

The fourth objective is that n be sufficient to ensure, with probability $1 - \beta$, that a difference in estimated proportions, $\hat{p}_1 - \hat{p}_2 > 0$, will be statistically significant at a specified type-I error rate (α, typically $\alpha = 0.05$) if in fact $p_1 - p_2 > 0$. This probability, $1 - \beta$, is also known as the "power" of the statistical test. This objective is relevant to the statistical comparison of proportions between strata or substrata (e.g., nearshore versus offshore) for a given stock, or between stocks within a stratum. Since the frame of reference for such hypotheses is typically the ocean mixed stock aggregate rather than the fishery landings, we assume independent binomial sampling for each of the two proportions, and assume equal sample size, $n_1 = n_2 = n$. We used the approximation provided by Casagrande et al. (1978) for the n required in this setting,

$$
n = \frac{A \left(1 + \sqrt{1 + \frac{4d}{A}} \right)^2}{4d^2},
\tag{14}
$$

where $A = \left(Z_{1-\alpha} \cdot \sqrt{2\overline{p}(1-\overline{p})} + Z_{1-\beta} \cdot \sqrt{p_1(1-p_1) + p_2(1-p_2)} \right)^2$, $d = p_1 - p_2$, and $\overline{p} = \frac{p_1 + p_2}{2}$.

Figure 4 shows the n necessary to yield a specified power, depending on p_1 (the larger proportion), for select values of the ratio $\frac{p_2}{p_1}$. Mohr (2007) suggests that the power be at least 0.8 for all $p_1 \geq 0.03$, assuming $\frac{p_2}{p_1} \leq 0.5$ and $\alpha = 0.05$. Since for any particular value of power the required n increases as p_1 approaches 0, and for fixed p_1 as $\frac{p_2}{p_1}$ approaches 1, the minimum n that will meet this condition is that which satisfies an Objective 4 criterion of power = 0.8 and $\alpha = 0.05$ for
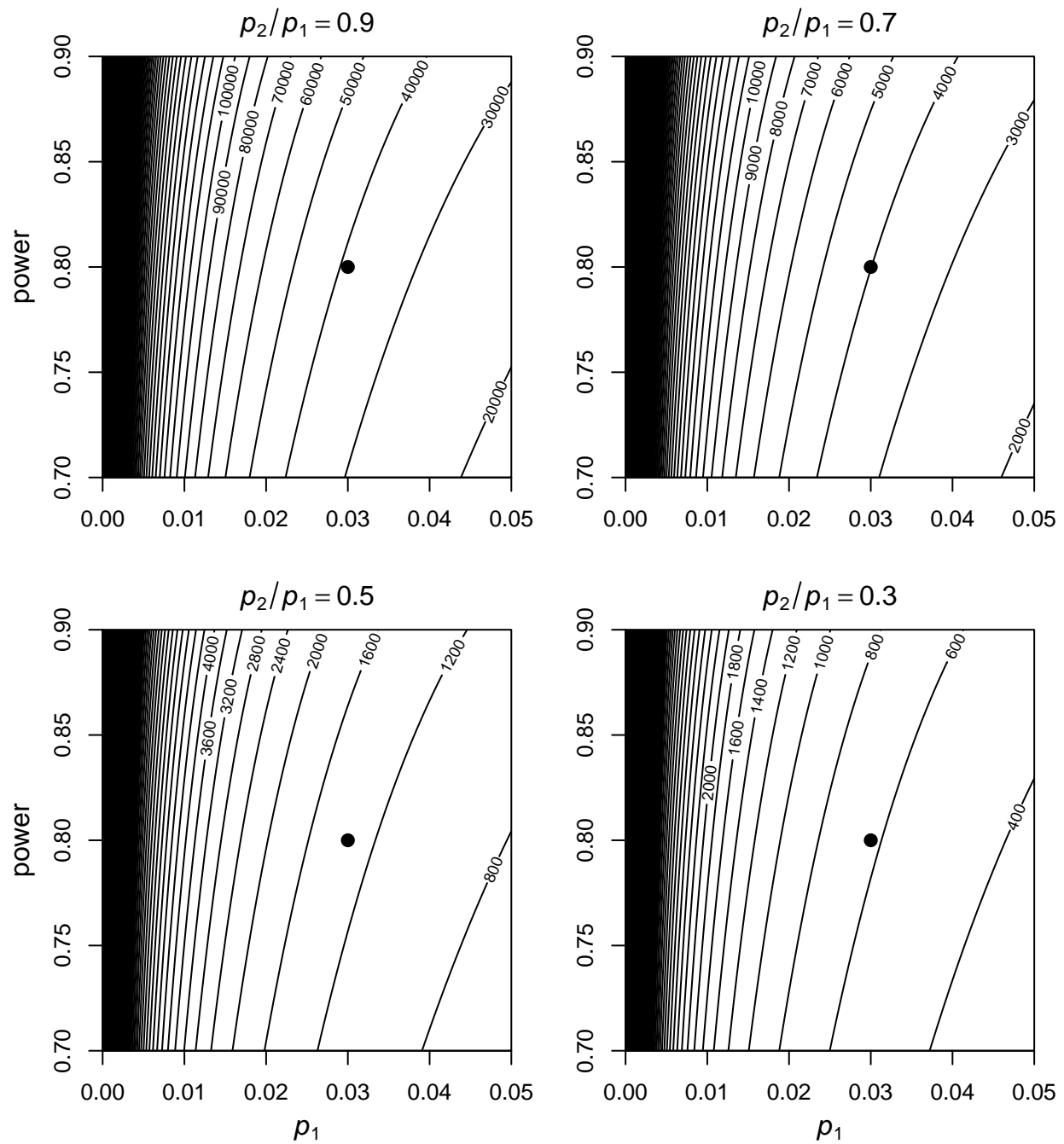
Figure 4. Sample size (contours) required to yield a specified power (with $\alpha = 0.05$) to detect a difference in two stock proportions, given the larger proportion p_1, and the ratio of the two proportions $\frac{p_2}{p_1}$ (different panels). Sample size indicated is required for each of the two independent binomial samples. Large dot marks criterion described in the text (power $= 0.8$ for $p_1 = 0.03$).

$p_1 = 0.03$ with $\frac{p_2}{p_1} = 0.5$. This is achieved with an n of 1338 fish per stratum or substratum being compared.

As with Objectives 1 and 2, the required n scales with the inverse of the stock proportions being compared when these proportions are small. Comparison of the panels in Figure 4 reveals large increases in the required n as the stock proportions being compared become more similar, compared to relatively modest changes in the required n over the range of power explored for a given $(p_1, p_2 = p_1 \cdot \frac{p_2}{p_1})$. For example, for the criterion above (power = 0.8 and $\alpha = 0.05$ for $p_1 = 0.03$), with $\frac{p_2}{p_1} = (0.3, 0.5, 0.7, 0.9)$ the required n is (627, 1338, 4012, 38703), whereas the n required for power = (0.7, 0.9) is approximately (80%, 135%) of that required for power = 0.8, and this scaling varies little as $\frac{p_2}{p_1}$ ranges between 0 and 1.

5   Discussion

The sample size requirements for any sampling program depend on the specific goals and questions the sampling is intended to address, as well as situation-specific constraints. Nevertheless, the results presented in this paper can provide some guidance on the general utility of current and proposed sampling schemes, as well as some sense of the potential benefits of increased sampling and what is sacrificed by decreased sampling. The West Coast Salmon GSI Collaboration stated a goal of successfully sampling 800 Chinook salmon from each fishery management area in California and Oregon per month (Goldenberg and Fitzpatrick 2010). Thus, the sampling target of 800 fish per stratum is a reasonable starting point for discussion.

If the intent is to estimate the contribution of a single stock to the ocean mixed stock aggregate, sampling 800 fish per stratum will have $> 99.9\%$ probability of detecting a stock with p as low as 0.01, allow estimation of proportions as small as 0.03 with $CV_{\hat{p}} \leq 0.2$, and yield a maximum $\sigma_{\hat{p}}$ of 0.018 for $p = 0.5$. Sampling 800 fish per stratum or substratum will also yield a power of at least 80% (at $\alpha = 0.05$) to detect differences between proportions, the smaller of which is no more than half of the larger, when the larger proportion is ≥ 0.05. Sample size requirements for comparisons

become much larger as the proportions become more similar. Thus, while a sample size of 800 fish per stratum suffices to answer some interesting questions, it should be noted that strong inference about rare stocks or about differences in their stock proportions requires more intensive sampling. For example, a sample size of 800 fish would result in a $CV_{\hat{p}}$ of 0.5 for $p = 0.005$, worse yet for smaller p. Below, we offer a rule of thumb to determine how much the sample size must be increased to meet specific goals.

Objective 3 is unique among the set of objectives in that it depends on absolute rather than relative precision, and as a result is more difficult to achieve for $p = 0.5$ rather than for small (or large) p. Thus, Objective 3 may have limited relevance to studies where rare stocks are of interest. For Objective 3, with any particular p, the required sample size simply scales with the inverse of the square of the desired precision—that is, to halve the standard error the sample size must be quadrupled. This also holds true for relative precision—for any particular p, the sample size must be quadrupled to halve the $CV_{\hat{p}}$.

The remaining objectives become increasingly difficult to achieve as p becomes small. We are able to offer a simple rule of thumb that approximates the potential benefits of increasing sample size beyond the current standard of 800 fish per stratum. For these objectives, a "small p" approximation can be applied where, once a sample size n adequate for a small p has been determined, the sample size needed to reliably estimate an even smaller proportion, p', can be readily approximated as $n \cdot \frac{p}{p'}$. For example, estimating a target proportion half as large as the currently achievable objective requires twice the sample size. This provides context for predicting the anticipated benefits of increased sampling effort. For instance, one could determine whether sampling 1600 rather than 800 fish justifies the extra expense by considering the benefits of reducing the minimum estimable proportion from 0.03 to 0.015. Note however that these "small p" approximations only apply within the range of sample sizes and target proportions where p is already "small". Thus, a similar approach would underestimate the lost precision resulting from reducing the sample size from 800 to 400, more so for further reductions. For Objectives 1 and 2, the "small p" approximation applies well for anticipating the benefits of increased sampling beyond 800 fish, but for

Objective 4 the "small p" approximation requires $p_1 \leq 0.02$ and does not apply closely unless all sample sizes under consideration are on the order of 2000 or more.

Thus far, our discussion has centered on the stock composition of the ocean mixed stock aggregate. If, instead, interest is restricted to the stock composition of the landings itself, smaller sample sizes will suffice. Note however that in this case, obtaining performance equivalent to what $n = 800$ provides under binomial sampling may require sampling a large fraction of the landings. For example, similar performance with respect to Objectives 1–3 would require sampling 44–55% of the landings for $N = 1000$, and 62–80% of the landings for $N = 500$.

Note also that all sample size recommendations in this paper refer to the number of fish successfully processed and genotyped. Larger sample collections may be required to account for potential sample loss, poor tissue preservation, or problems with the genotyping of individual samples. An additional caveat to the presented sample size targets is that our analysis treats each sampled fish as having been independently selected in a "simple random" fashion. For the ocean mixed stock aggregate assessment, this assumption is likely violated by the propensity of fish to school (Burnham and Anderson 1998) and because fishermen tend to fish nonrandomly. Required sample sizes in the presence of such violations are larger than the sample sizes presented herein. At-sea sampling programs would be expected to violate these assumptions to a larger degree than dock-side sampling programs, since fewer vessels would be providing the samples, and sampling fractions for these vessels would likely be higher (or even 100%), exacerbating the likely non-independence of the individual fish sampled.

A further planning consideration for GSI sampling is that in many cases the stock units of interest may be a subset of a genetic reporting group. For example, one may need to be able to estimate the age-specific proportions of a specific population or source river within a particular reporting group. This finer-scale resolution would be required to use GSI-derived data as an input for cohort reconstruction, for example (O'Farrell et al. 2012). Thus the stock proportions of interest for management may be substantially smaller than the stock proportions estimated for genetically-defined reporting groups. And of course, such estimates would only be possible if individual fish

were aged as well as genotyped.

This paper has considered estimation of a single stock proportion at a time (or two proportions for Objective 4, although the proportions may be from distinct strata). Bernard (2007), Bromaghin (1993), and Thompson (1987) address sample size requirements for a multinomial sampling model. Required sample sizes for precise simultaneous estimation of multiple proportions may be larger than required for estimating a single proportion.

We also note that, while estimates of stock composition based on GSI data may be made with more confidence using mixed stock analysis than with individual fish assignments (Koljonen et al. 2005), an approach based on individual-specific assignments would seem to be necessary when incorporating auxiliary information such as individual size, age, or fine-scale catch location data.

Finally, we have not addressed the impacts of genetic assignment error in this paper. The effects of genetic assignment errors on the estimated stock composition are complex. Classification accuracy varies by stock, but is generally fairly high at the reporting group level of aggregation (Anderson et al. 2008; Seeb et al. 2007). Nevertheless, low misclassification rates can still result in biased estimates of stock proportions, particularly when abundant stocks are misidentified as rare stocks. The magnitude of this bias depends on the resolution of the genetic baseline and the proportions of fish in the mixture. A more thorough exploration of the implications of assignment error for sample size requirements, or the potential biases introduced by stock-dependent misassignment probabilities, is beyond the scope of this paper.

6 Acknowledgements

References

Anderson, E. C., R. S. Waples, and S. T. Kalinowski (2008). An improved method for predicting the accuracy of genetic stock identification. *Canadian Journal of Fisheries and Aquatic Sciences 65*(7), 1475–1486.

Bernard, D. R. (2007). Catch sampling to determine stock composition with GSI: sample sizes and procedures. In J. Korman, G. Morishima, and D. Hankin (Eds.), *Report of Discussions from GSI Modelling and Sampling Workgroup (Portland, OR, May 16-17, 2007), and Compilation of Workshop Reports Presented at the September 11-13, 2007 Workshop (Vancouver)*, pp. 103–108.

Bromaghin, J. F. (1993). Sample size determination for interval estimation of multinomial probabilities. *The American Statistician 47*(3), 203–206.

Burnham, K. P. and D. R. Anderson (1998). *Model selection and inference: a practical information-theoretic approach*. New York: Springer Verlag.

Casagrande, J. T., M. C. Pike, and P. G. Smith (1978). An improved approximate formula for calculating sample sizes for comparing two binomial distributions. *Biometrics 34*(3), 483–486.

Cochran, W. G. (1977). *Sampling techniques, 3rd edition*. New York: Wiley.

Fosgate, G. T. (2009). Practical sample size calculations for surveillance and diagnostic investigations. *Journal of Veterinary Diagnostic Investigation 21*(1), 3–14.

Goldenberg, D. and N. Fitzpatrick (2010). The West Coast Salmon Genetic Stock Identification Collaboration Annual Activity Report, 2010. Submitted to Pacific Fishery Management Council, March 2011. Available from `http://www.pcouncil.org/wp-content/uploads/G6b_GSI_RPT_MAR2011BB.pdf`.

Goldwasser, L., M. S. Mohr, A. M. Grover, and M. L. Palmer-Zwahlen (2001). The supporting databases and biological analyses for the revision of the Klamath Ocean Harvest Model. Unpublished report. National Marine Fisheries Service, Santa Cruz, CA. Available from M.S. Mohr, National Marine Fisheries Service, 110 Shaffer Road, Santa Cruz, CA 95060, USA.

Johnson, J. K. (1990). Regional overview of coded wire tagging of anadromous salmon and steel-

head in northwest America. *American Fisheries Society Symposium 7*, 782–816.

Johnson, N. L., S. Kotz, and A. W. Kemp (1993). *Univariate discrete distributions, 2nd edition*. New York: Wiley.

Koljonen, M.-L., J. J. Pella, and M. Masusda (2005). Classical individual assignments versus mixture modeling to estimate stock proportions in Atlantic salmon (*Salmo salar*) catches from DNA microsatellite data. *Canadian Journal of Fisheries and Aquatic Sciences 62*(9), 2143–2158.

Mohr, M. (2007). Sample size considerations for GSI studies. In J. Korman, G. Morishima, and D. Hankin (Eds.), *Report of Discussions from GSI Modelling and Sampling Workgroup (Portland, OR, May 16-17, 2007), and Compilation of Workshop Reports Presented at the September 11-13, 2007 Workshop (Vancouver)*, pp. 95–102.

O'Farrell, M. R., S. D. Allen, and M. S. Mohr (2012). The Winter-Run Harvest Model (WRHM). *NOAA Technical Memorandum NOAA-TM-NMFS-SWFSC-489*.

O'Farrell, M. R., W. H. Satterthwaite, and B. C. Spence (2012). California Coastal Chinook salmon: status, data, and feasibility of alternative fishery management strategies. *NOAA Technical Memorandum NOAA-TM-NMFS-SWFSC-494*.

OSP (Ocean Salmon Project) (1999). 1998 ocean salmon fisheries annual report. California Department of Fish and Game. Available from `http://docs.streamnetlibrary.org/StreamNet_References/CAsn90419.pdf`.

Satterthwaite, W. H., M. S. Mohr, M. R. O'Farrell, and B. K. Wells (2012). A Bayesian hierarchical model of size-at-age in ocean-harvested stocks – quantifying effects of climate and temporal variability. *Canadian Journal of Fisheries and Aquatic Sciences 69*(5), 942–954.

Satterthwaite, W. H., M. S. Mohr, M. R. O'Farrell, and B. K. Wells (2013). A comparison of temporal patterns in the ocean spatial distribution of California's Central Valley Chinook salmon runs. *Canadian Journal of Fisheries and Aquatic Sciences 70*(4), 574–584.

Seeb, L. W., A. Antonovich, M. A. Banks, T. D. Beacham, M. R. Bellinger, S. M. Blankenship,

M. R. Campbell, N. A. Decovich, J. C. Garza, C. M. Guthiree III, T. A. Lundrigan, P. Moran, S. R. Narum, J. J. Stephenson, K. J. Supernault, D. J. Teel, W. D. Templin, J. K. Wenburg, S. F. Young, and C. T. Smith (2007). Development of a standardized DNA database for Chinook salmon. *Fisheries 32*(11), 540–552.

STT (Salmon Technical Team) (2000). STT recommendations for hooking mortality rates in 2000 recreational ocean Chinook and coho fisheries. Pacific Fishery Management Council, 7700 NE Ambassador Place, Suite 101, Portland, Oregon 97220-1384. Available from `http://pcouncil.org/wp-content/uploads/sttb20302.pdf`.

Thompson, S. K. (1987). Sample size for estimating multinomial proportions. *The American Statistician 41*(1), 42–46.

Trudel, M., J. Fisher, J. A. Orsi, J. F. T. Morris, M. E. Thiess, R. M. Sweeting, S. Hinton, E. A. Fergusson, and D. W. Welch (2009). Distribution and migration of juvenile Chinook salmon derived from coded wire tag recoveries along the continental shelf of western North America. *Transactions of the American Fisheries Society 138*(6), 1369–1391.

Weitkamp, L. and K. Neely (2002). Coho salmon (*Oncorhynchus kisutch*) ocean migration patterns: insight from marine coded-wire tag recoveries. *Canadian Journal of Fisheries and Aquatic Sciences 59*(7), 1100–1115.

Weitkamp, L. A. (2010). Marine distributions of Chinook salmon from the west coast of North America determined by Coded Wire Tag recoveries. *Transactions of the American Fisheries Society 139*(1), 147–170.

Winans, G. A., D. Viele, A. Grover, M. Palmer-Zwahlen, D. Teel, and D. Van Doornik (2001). An update of Genetic Stock Identification of Chinook salmon in the Pacific Northwest: test fisheries in California. *Reviews in Fisheries Science 9*(4), 213–237.

Appendix A   Evaluating Objective 1 for $k > 1$

Objective 1 is that n be sufficient to ensure, with probability Q_k, that the sample will contain at least k fish of the stock of interest. That is, $Q_k = \text{Prob}(Y \geq k) = 1 - \text{Prob}(Y < k)$. Here, we evaluate this objective for $k > 1$; the case of simple detection ($k = 1$) is evaluated in Section 4.1 of the main text. For binomial sampling, appealing to equation (1)

$$Q_k = 1 - \sum_{i=0}^{i=k-1} \text{Prob}(Y = i) = 1 - \sum_{i=0}^{i=k-1} \binom{n}{i} p^i (1-p)^{n-i}. \tag{A-1}$$

For hypergeometric sampling, appealing to equation (3),

$$Q_k = 1 - \sum_{i=0}^{i=k-1} \text{Prob}(Y = i) = 1 - \sum_{i=0}^{i=k-1} \frac{\binom{Np}{i}\binom{N(1-p)}{n-i}}{\binom{N}{n}}. \tag{A-2}$$

An analytical solution of these expressions in terms of n is not available for $k > 1$, but the required sample size may be found by searching numerically for the smallest n that satisfies or exceeds the desired level of Q_k using equation (A-1) or (A-2), respectively. Figures A-1 and A-2 show the n necessary to yield a specified Q_5 or Q_{10}, respectively, depending on p, under both binomial sampling and hypergeometric sampling for select values of N. Clearly, higher values of k require higher n, for all values of p and Q_k. For binomial sampling, a criterion of $Q_5 = 0.999$ for $p = 0.01$ requires an n of 1475 (Figure A-1), while a criterion of $Q_{10} = 0.999$ for $p = 0.01$ requires an n of 2259 (Figure A-2). This illustrates that, for fixed Q_k, doubling k does not necessarily require doubling n. However, for $Q_k = 0.999$, halving p from 0.01 to 0.005 does require an approximate doubling of n for a particular value of k; for $k = 5$, $n = 2954$, and for $k = 10$, $n = 4525$.

For hypergeometric sampling, as N is reduced so too is the required n, with n approaching N (100% sampling) for small p. For $N = 1000$, binomial and hypergeometric sampling require similar n for $p \geq 0.02$, but as p approaches 0 the n required under hypergeometric sampling increases at a much slower rate than for binomial sampling. No contours appear in the Figure A-1 and A-2 $N = 100$ panels because, in this case, the minimum possible value of p associated with Q_k, $p = \frac{k}{N}$,

is $p = 0.05$ and $p = 0.10$ for Q_5 and Q_{10}, respectively, which is outside the plotted range of p (except for Q_5 with $p = 0.05$ where $Q_5 = 1$ if $n = N = 100$).
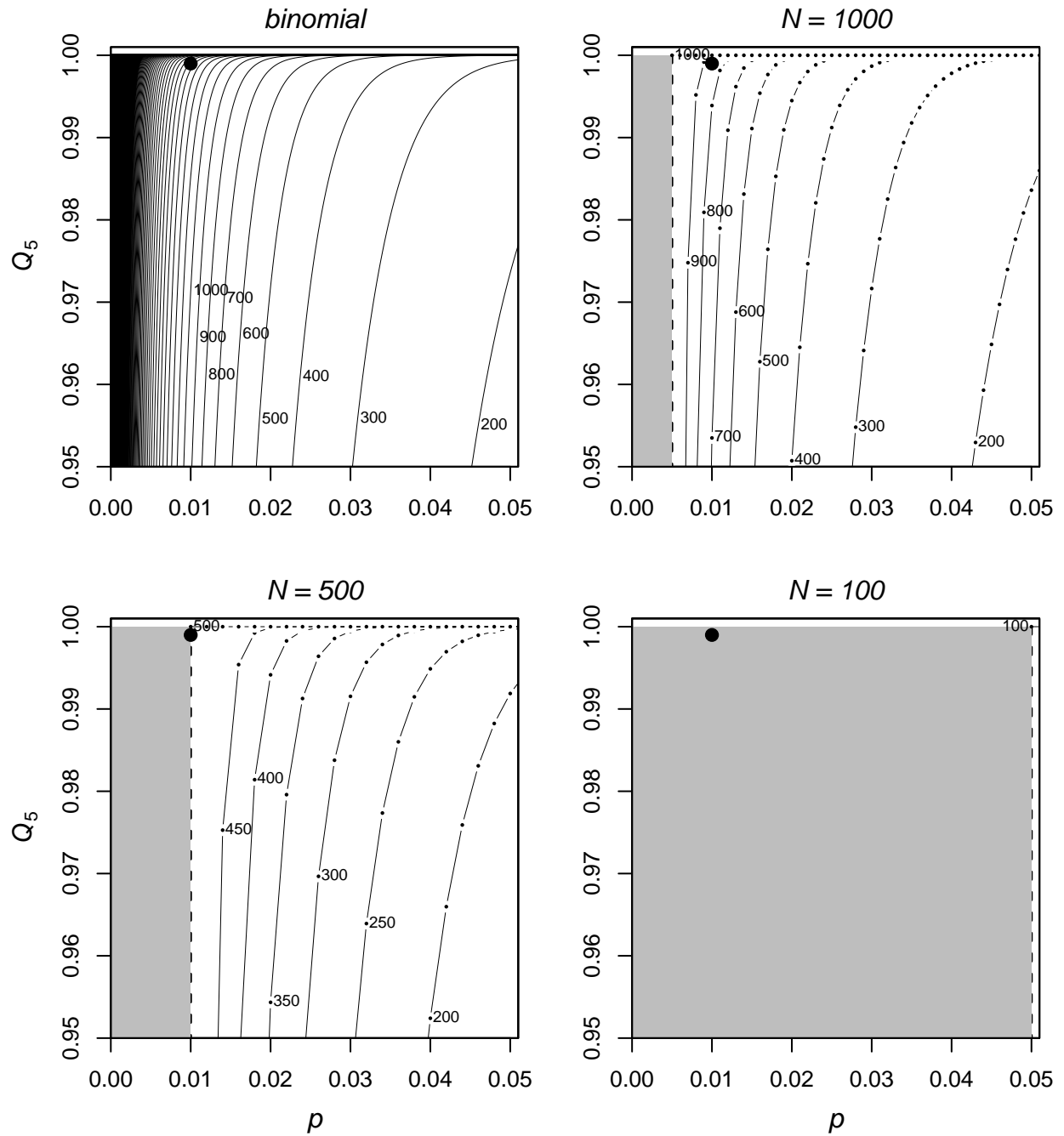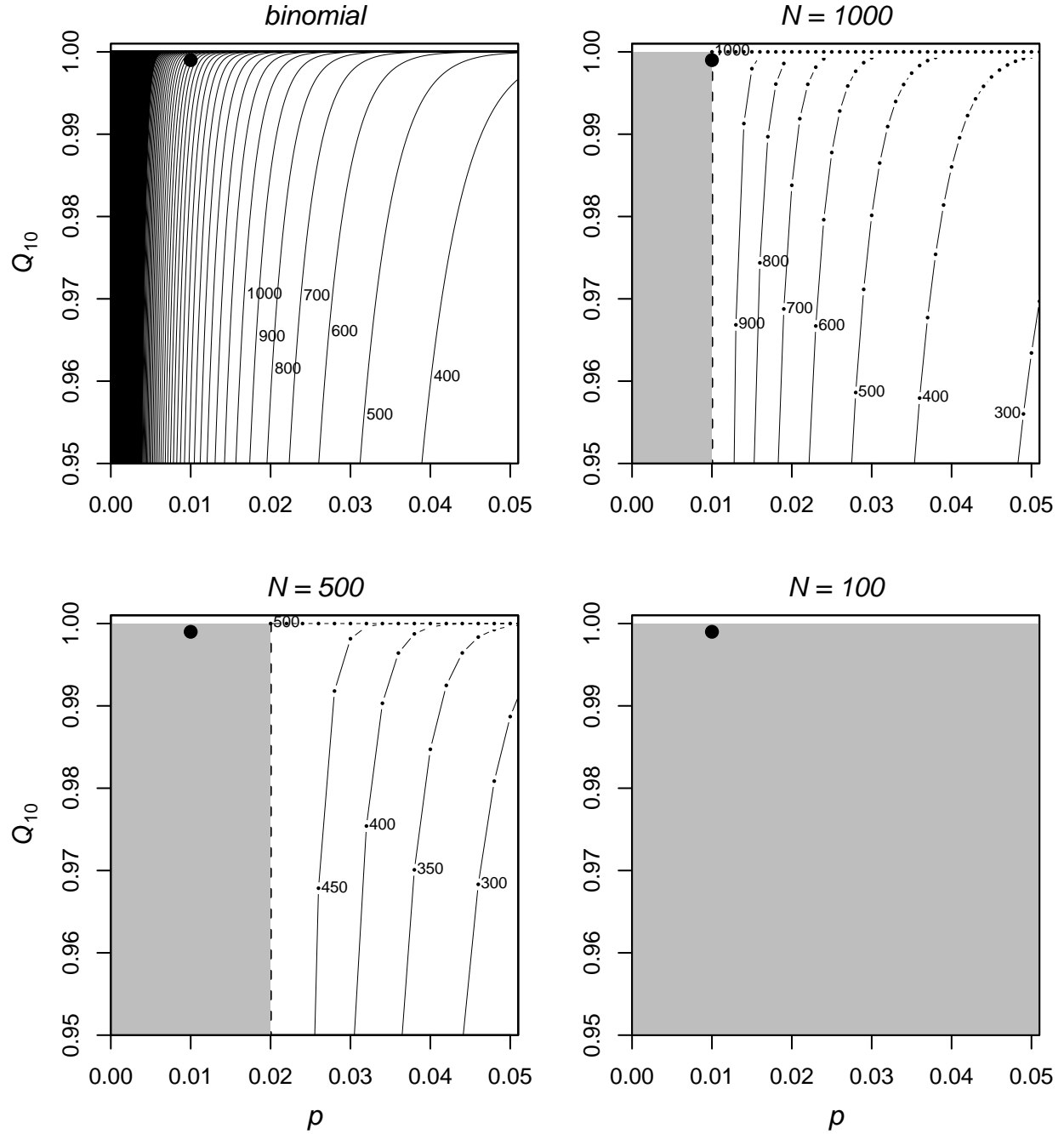
Figure A-1. Sample size (contours) required to ensure, with probability Q_5, that the sample will contain at least five fish of the stock of interest, given the stock proportion p. Top-left panel presents results for binomial sampling; remaining panels give hypergeometric results for total landings of 1000, 500, and 100 fish, respectively. In the hypergeometric panels, p less than the minimum value possible $\left(\frac{5}{N}\right)$ for Q_5 are block-shaded in grey, and for contoured values of n the possible (n, p) combinations appear as black dots connected by line segments. Large dot marks criterion described in the text ($Q_5 = 0.999$ for $p = 0.01$). Note that for $N = 100$ this criterion cannot be satisfied, even for $n = N$.

Figure A-2. Sample size (contours) required to ensure, with probability Q_{10}, that the sample will contain at least ten fish of the stock of interest, given the stock proportion p. Top-left panel presents results for binomial sampling; remaining panels give hypergeometric results for total landings of 1000, 500, and 100 fish, respectively. In the hypergeometric panels, p less than the minimum value possible $\left(\frac{10}{N}\right)$ for Q_{10} are block-shaded in grey, and for contoured values of n the possible (n, p) combinations appear as black dots connected by line segments. Large dot marks criterion described in the text ($Q_{10} = 0.999$ for $p = 0.01$). Note that for $N = 500$ and $N = 100$ this criterion cannot be satisfied, even for $n = N$.

RECENT TECHNICAL MEMORANDUMS